

1 WORD COUNT: 8,550

2 WORD COUNT (including references): 11,350

3 **Models linking production and comprehension**

4 Chiara Gambi

5 Saarland University and University of Edinburgh

6 Martin J. Pickering

7 University of Edinburgh

8

9 Address for correspondence:

10 Chiara Gambi

11 Developmental Lab

12 Department of Psychology

13 University of Edinburgh 7 George Square

14 Edinburgh EH8 9JZ

15 [cgambi@exseed.ed.ac.uk](mailto:cgambi@exseed.ed.ac.uk); [g.chiara86@gmail.com](mailto:g.chiara86@gmail.com)

16 Phone: +44 (0)131 650 3343

17

18

19

20

21

22

23

# Models linking production and comprehension

## Introduction

Production and comprehension have traditionally been studied in separate subfields of psycholinguistics, despite the fact that many psycholinguistic tasks involve both. For example, in the picture-word interference paradigm, participants name target pictures while ignoring auditorily presented or superimposed written distractor words (e.g., Schriefers, Meyer, & Levelt, 1990). Therefore, this paradigm measures the effect of comprehension processes on production processes. This disciplinary division notwithstanding, there is general consensus that production and comprehension are linked. However, opinions diverge with regard to the nature of this link. It is not easy to characterise scholarly disagreement on this issue, as the issue is often not discussed explicitly, and even when it is, it is rarely treated as a key element of the model being proposed.

There are, however, some exceptions to this trend. Below, we focus on models that make explicit claims about the links between production and comprehension. It is noteworthy that these models tend to emphasize what is shared between comprehension and production, rather than what is not shared. Not surprisingly, frameworks that emphasize sharing tend to be concerned with explaining language learning, acquisition, and change (see Dell & Chang, 2014; MacDonald, 2013) or dialogue (see Pickering & Garrod, 2013), rather than isolated acts of production or comprehension. In fact, the simple observation that the primary site of language use is dialogue constitutes, in itself, a strong motivation for positing links between production and comprehension (Garrod & Pickering, 2004; see section on *Dialogue*, this chapter). And theories of how linguistic representations are acquired and develop over time must naturally take into account how linguistic input shapes the output of the language system (and *vice versa*; see *Frameworks that posit linked preferences*, this chapter).

Traditionally, a number of arguments are used to claim that a separation between comprehension and production should be maintained. First, dissociations between the comprehension and production abilities of patients with brain lesions appear to support the notion that the neural substrates of production and comprehension are separate. For example, Kim and Thompson (2000) reported that agrammatic patients showed intact comprehension of verbs but were impaired in verb naming. Second, asymmetries exist between the development of language comprehension and language production in infants (see Bates, 1993), and similarly, between the rate of decay of comprehension and production abilities in older adults (e.g., McKay, Abrams, & Pedroza, 1999). These findings indicate that it is not possible to fully equate production and

comprehension. A comprehensive review of these and other arguments is beyond the scope of this chapter. Here, we just note that these arguments do not necessarily imply that comprehension and production are completely separate. Rather, dissociations and asymmetries are in principle compatible with some degree of sharing, as long as there are subcomponents of production that are not used in comprehension and *vice versa*.

When discussing the issue of what is shared between production and comprehension, it is useful to bear in mind the distinction between linguistic *representations* and *processes* acting on those representations. Linguistic representations are the components of memory that store information about linguistic units (e.g., phonemes, words, syntactic rules, concepts). Comprehension and production processes are the cognitive operations that can be applied to linguistic representations (e.g., retrieval, spreading activation, inhibition), as well as the operations that map from abstract representations to articulation and from acoustics to abstract representations. Processes are directional: so, for example, the process that retrieves a phonological representation given an activated semantic representation is not the same as the process that retrieves a semantic representation given an activated phonological representation. Below, we first discuss sharing with regard to representations (*Representational parity*), and then we turn to theories that posit common processes (*Linked processes*).

## **Representational parity**

Most theorists assume that some of the representations that are accessed during production are the same as the representations accessed during comprehension. At the single word level, the most influential theory of lexical access in production assumes conceptual (semantic), lemma (syntactic), and word-form (sound-based) representations, and proposes the network for production and the network for comprehension “coincide from the lemma level upwards” (i.e., concept and lemma nodes are shared between production and comprehension; Levelt, Roelofs, & Meyer, 1999; p. 7). Pickering and Garrod (2004) also assume shared representations (which they term the *parity hypothesis*), but extend the scope of this assumption by positing that representations used in comprehension are the same as representations used in production at all linguistic levels, from the situation model (i.e., a representation of the situation being discussed, including time, space, causal relations, intentionality, and individuals involved; Zwaan & Radvansky, 1998) to phonology and phonetics. Shared phonological representations are also part of the Node Structure Theory of MacKay (1987). Moreover, parity at the sound level is the central assumption of the Motor Theory

1 of speech perception (Lieberman & Whalen, 2000; see Galantucci, Fowler, & Turvey, 2006), and of  
2 the Episodic Theory of speech perception (Goldinger, 1998).

3 Parity at the semantic and lexico-syntactic levels has been confirmed by several studies. The  
4 strongest evidence comes from findings of immediate effects of comprehension on production.  
5 First, silent reading of words semantically and associatively related to the name of a target picture  
6 affects naming times for the picture (e.g., Schriefers, et al., 1990; Alario, Segui, & Ferrand, 2000).  
7 Second, silent reading of sentences such as *A rock star sold an undercover agent some cocaine*  
8 (double object, DO) or *A rock star sold some cocaine to an undercover agent* (prepositional object,  
9 PO) influences what sentence structure (DO or PO) is used to describe a target scene (depicting an  
10 unrelated event, such as a man reading a book to a boy; Bock, Dell, Chang, & Onishi, 2007). And,  
11 similarly, the syntactic choices of their interlocutor influence affect speakers' syntactic choices in  
12 dialogue (Branigan, Pickering, & Cleland, 2000; Levelt & Kelter, 1982; see Pickering & Ferreira,  
13 2008 for a review). In addition, interlocutors align their lexical choices (Brennan & Clark, 1996;  
14 Garrod & Anderson, 1987).

15 As well as behavioural evidence, there is growing evidence for lexico-syntactic parity at the  
16 neural level. Two fMRI studies showed that the same neural populations are recruited during  
17 comprehension and production of sentences (Menenti, Gierhan, Segaert, & Hagoort, 2011; Segaert,  
18 Menenti, Weber, Petersson, & Hagoort, 2012). They identified brain areas that show decreased  
19 activation while producing (or comprehending) a given sentence structure, when the participant had  
20 just processed a sentence with the same structure, compared to a sentence with a different structure.  
21 This phenomenon is called repetition suppression, and it is used to localize neural areas that are  
22 sensitive to a given property of the stimulus (in this case, structure). Importantly, the areas  
23 identified were the same regardless of whether prior processing of the same structure had taken  
24 place in production or comprehension.

25 Regarding parity at the phonological level, it is known that silent reading of, or passive  
26 listening to, distractor words phonologically related to the name of a target picture speeds up  
27 naming times for the picture (e.g., Schriefers, et al., 1990; Damian & Martin, 1999). This is usually  
28 taken as evidence that comprehension of distractor words pre-activates phonological representations  
29 they share with the target, so that those representations are subsequently easier to access in  
30 production. Moreover, Kerzel and Bekkering (2000) found that participants pronounced a printed  
31 syllable while watching a video of a mouth producing the same syllable more quickly than when the  
32 mouth produced a different syllable (see also Jarick & Jones, 2008). In addition, Galantucci,  
33 Fowler, and Goldstein (2009) showed that speakers are faster to produce a syllable (e.g., /ba/) if

they have just listened to the same syllable than to a syllable with a different onset (e.g., /da/) (see also Fowler, Brown, Sabadini, & Weihing, 2003).

There is also much evidence demonstrating motor activation during the perception of speech. Such evidence suggests that speech production and speech perception make use of overlapping neural representations. Several studies have found activation of motor areas in the brain during audiovisual speech perception (e.g., Skipper, Nusbaum, & Small, 2005; Skipper, van Wassenhove, Nusbaum, & Small, 2007), and also during passive listening to speech (e.g., Wilson, Saygin, Sereno, & Iacoboni, 2004). Importantly, motor activation during passive listening is articulator-specific: for example, listening to labial consonants is associated with activation of the lip representation area in motor cortex (Pulvermüller et al., 2006; see Pulvermüller & Fadiga, 2010 for a review). Further, listening to speech modulates the excitability of the speech muscles that are involved in the production of the perceived sound (e.g., Fadiga, Craighero, Buccino, & Rizzolatti, 2002). Accordingly, listening to a phoneme also affects concurrent articulation of a different phoneme; for example the palatal sound /k/ was produced with greater contact between the tip of the tongue and the alveolar ridge when participants were listening to the alveolar sound /t/ (Yuen, Davis, Brysbaert, & Rastle, 2010).

This body of evidence supports the assumption of shared representations at the phonological level at least. However, recent findings suggest that motor activation during speech perception might occur primarily or exclusively when the task is difficult (e.g., when speech is degraded; Adank, 2012; D’Ausilio, Bufalari, Salmas, & Fadiga, 2012). Moreover, it may be that motor activation under normal listening conditions reflects listeners’ tracking of the speaker’s speech rate and preparation for speech in anticipation of the end of the speaker’s utterance, rather than retrieval of shared content-specific phonological representations (S. K. Scott, McGettigan, & Eisner, 2009).

INSERT FIGURE 1ABOUT HERE

Figure 1A is intended as a schematic summary of the overview of the literature on representational parity presented in this section. It depicts what we take to be the consensus view on the issue of representational parity across levels of linguistic representations. First of all, it illustrates the fact that there is substantial consensus on parity at the semantic and syntactic level (grey *sem* and *syn* representations). Phonological representations, instead, are depicted as

overlapping but not identical (partially superimposed black and white *phon* representations) to account for the fact that evidence for parity at this level might be restricted to particular task demands (as just discussed). Finally, in Figure 1 representations at the phonetic level are labelled as speech percepts (comprehension) and speech motor commands (production) and are assumed to be separate (i.e., white speech percepts are separate from black motor command representations).

Some researchers assume substantial overlap at the phonetic level as well, but parity at this level is particularly controversial. On one hand, speakers can converge towards a model speaker at the level of low-level phonetic features that do not imply phonological distinctions (e.g., vowel duration, F0), therefore providing support for the hypothesis that phonetic representations are shared between comprehension and production (e.g., Goldinger, 1998; Pardo, 2006; see Chapter 6 in this volume). However, other studies failed to find evidence that listening to a phonetic variant (e.g., alveolar /r/) facilitates subsequent production of the same variant compared to listening to an alternative variant (e.g., uvular /r/) that is phonologically equivalent (Mitterer & Ernestus, 2008; Mitterer & Müsseler, 2013).

Moreover, despite some indication that adaptation to an accent in comprehension correlates with adaptation to the same accent in production at the level of an individual speaker (Evans & Iverson, 2007), there are also asymmetries between people's production and comprehension abilities in the processing of regional variability in speech. For example, Sumner and Samuel (2009) argued that listeners who have long-term perceptual experience with a dialectal phonetic variant can achieve native-like perception of that variant despite lacking native-like production representations (see also Kralijc, Brennan, & Samuel, 2008). To sum up, some findings in the literature suggest comprehension and production phonetic representations are shared or overlapping, but others support the notion of distinct comprehension and production representations at this level. Overall, the evidence for separation is stronger at the phonetic than at the phonological level. In Figure 1A, we capture this by drawing separate (rather than partially overlapping) phonetic representations, but it is still unclear how much separation should be assumed at this level.

## **Linked processes**

Figure 1 does not only depict representations, but also processes. Traditionally, any process that takes place during an act of comprehension (i.e., while reading or listening) has been considered a comprehension process and, conversely, any process that takes place during an act of production (i.e., while writing or speaking), a production process. Within this tradition, the debate

1 has focussed on whether processes that take place during production always map from semantics to  
2 syntax and from syntax to phonology (i.e., white arrows on black background, flowing from left to  
3 right in Figure 1A), or whether they can also map in the opposite direction (i.e., black arrows on  
4 white background, flowing from right to left in Figure 1A). Processes that go “backwards” have  
5 been termed feedback processes and models that incorporate them (e.g., Dell, 1986) have been  
6 labelled as interactive (as opposed to purely feed-forward models; Levelt, et al., 1999). Similar  
7 issues have been discussed in the comprehension literature, where left-to-right processes (see Figure  
8 1A) have been called top-down and right-to-left processes have been called bottom-up (e.g.,  
9 Marslen-Wilson, 1987). Within the traditional view, researchers do not typically ask whether  
10 production processes can take place during comprehension or whether comprehension processes  
11 can take place during production. Instead, they tend to label any process that takes place during  
12 production as a “production process” and any process that takes place during comprehension as a  
13 “comprehension process”, regardless of the direction in which it flows.

14 But on an alternative view, a process is classed as a production process if it maps from  
15 representations that are higher in the linguistic hierarchy (e.g., semantics) to representations that are  
16 lower (e.g., phonology), that is from left to right in Figure 1A. Conversely, a process is classed as a  
17 comprehension process if it maps from lower to higher representations (e.g., from phonology to  
18 semantics), that is from right to left in Figure 1A. According to this definition, which we will follow  
19 in the remainder of this chapter, both production and comprehension processes could potentially be  
20 employed during any act of production, as they could during any act of comprehension (Pickering  
21 & Garrod, 2013).

22 Note that this difference is not merely terminological. Rather, it reflects a substantial  
23 theoretical distinction between the traditional and the alternative view, which could be tested  
24 experimentally. Should we conceptualize a processing flow that goes from phonology to syntax  
25 during an act of production as a production process (because it takes place during an act of  
26 production, as per the traditional view), or as a comprehension process (because it flows from right  
27 to left, as per the alternative view)? One way of answering this question would be by examining the  
28 neural pathways involved. For example, if we could show that the same neural pathway is involved  
29 during an act of production that implies reliance on feedback flow of information, as well as during  
30 an act of comprehension, then we would have empirical evidence for the alternative view. But if  
31 feedback during production and comprehension engage separate pathways despite a common  
32 direction in the flow of information, then the evidence would be more compatible with the  
33 traditional view.

1           Unfortunately, progress towards establishing how much overlap there is between production  
2 and comprehension processes has been hindered by the division between sub-disciplines within  
3 psycholinguistics. Comprehension researchers often use a production measure (e.g., word naming  
4 time) as their dependent variable, but do not interpret the production processes themselves (the  
5 “mind-in-the-mouth” assumption; Bock, 1996). Similarly, production researchers presumably  
6 realize that a task such as picture-word interference involves comprehension processes but tend to  
7 ignore them.

8           For example, production researchers concerned with picture-word interference have  
9 assumed that word form and phoneme representations activated by distractor words (in  
10 comprehension) send activation to related word forms and phonemes in the production network;  
11 Levelt, et al., 1999). But they do not ask what processes are involved in comprehension of the  
12 distractor words, focussing instead only on the end result of those processes (i.e., that some  
13 representation gets activated). For example, Damian and Martin (1999) showed that the time course  
14 of semantic and phonological effects in picture-word interference differed for visually presented  
15 distractors compared to auditorily presented distractor words because of the longer presentation  
16 times used for visual distractors (in previous PWI studies). They did not consider the possibility that  
17 lexical co-activation (in comprehension) might have a different time course depending on modality  
18 (visual or auditory), and instead assumed that different presentation times led to different effects  
19 because the distractors interacted with different stages of production of the target picture name.

20           However, production-comprehension links have not been entirely overlooked. In fact, both  
21 feed-forward and feedback links have long been identified as a crucial component of neuro-  
22 computational theories of speech motor control (Tourville & Guenther, 2011; Hickok, 2012) and  
23 learning (Guenther & Vladusich, 2012; see also Plaut & Kello, 1999). Separately, the  
24 psycholinguistic literature on language production has also given ample consideration to this topic,  
25 albeit under the specific heading of self-monitoring. A long-standing psycholinguistic account of  
26 self-monitoring, the Perceptual Loop Theory, equates the self-monitoring system (active during acts  
27 of production) with the comprehension system (Levelt, 1983; 1989).

28           We first briefly present two neuro-computational models of speech motor control (Tourville  
29 & Guenther, 2011; Hickok, 2012). Then, we describe the Perceptual Loop Theory and discuss some  
30 criticisms of it. In the subsequent section, we introduce an integrated framework for language  
31 comprehension and language production (Pickering & Garrod, 2013), which includes an alternative  
32 account of self-monitoring (Pickering & Garrod, 2014), and also posits that production processes  
33 take place during acts of comprehension (and not just that comprehension processes take place



during acts of production, as in self-monitoring). We then consider the P-chain framework (Dell & Chang, 2014), which makes a related proposal. Finally, we discuss proposals that place the link between production and comprehension outside specific acts of comprehension or production, but instead in the long-term experience that speakers and listeners have with language (e.g., MacDonald, 2013).

### *Neuro-computational models of speech motor control*

In neuro-computational models of speech motor control (the Directions Into Velocities of the Articulators, or DIVA, model, see Tourville & Guenther, 2011; the Hierarchical State Feedback Control, or HSFC, model, see Hickok, 2012, 2014), forward models map from motor commands sent to the articulators to the sensory (i.e., auditory or somatosensory) consequences of executing those commands (the upward dotted vertical arrow from motor commands to speech percepts in Figure 1B). Forward models therefore instantiate a relatively low-level mapping between production and comprehension representations and can be considered as internal models of the language system (that is, models of the processes that cause the articulation of speech sounds during acts of production). The inverse mapping corresponds to feedback-based correction of speech movements (downward dotted vertical arrow from speech percepts to motor commands in Figure 1B).

Evidence that forward models are implicated in speech production comes from the finding that auditory responses to speech sounds are suppressed during speaking compared to listening (e.g., M100 suppression, as reported using magneto-encephalography in the study of Houde, Nagarajan, Sekihara, & Merzenich, 2002). This is thought to occur because forward models can be used during production to anticipate sensory stimulation and cancel it out (in a way that could be useful in distinguishing between self-generated and externally-generated sounds). In support of this claim, auditory responses are suppressed in a stimulus-specific manner: suppression occurs during covert rehearsal compared with a control task, but only when the rehearsed stimulus matches (part of) the perceived stimulus (Ylinen et al., 2014). In addition, enhancement rather than suppression takes place when auditory feedback is altered unexpectedly during speaking, for example by shifting pitch upwards or downwards in real-time, so that the predicted stimulation ceases to match actual stimulation (e.g., Chang, Niziolek, Knight, Nagarajan, & Houde, 2013).

All of these studies are compatible with forward-model predictions operating at the level of fine-grained phonetic features. But there is also some indication that such predictions might be

1 phonological in nature. Niziolek, Nagarjan, and Houde (2013) showed the degree of suppression is  
2 larger for sounds that are closer to a given speaker's median productions, suggesting that  
3 predictions could be computed on the basis of somewhat abstract representations. In other words,  
4 when the speaker selects a motor command to execute, the anticipated sensory consequences might  
5 correspond to an abstract phonological target (i.e., what it should sound like) rather than to a  
6 detailed phonetic target (i.e., what it is going to sound like on this particular instance).

7 This finding provides support for the HSFC model (Hickok 2012; 2014). In this model,  
8 forward model predictions operate at two hierarchically organized levels: phonemes and syllables.  
9 Motor programs (corresponding to planned syllables and planned phonemes) inhibit sensory areas  
10 where perceptual targets are represented. These same areas are activated via feedback from the  
11 movements of the vocal tract and from the resulting speech output (i.e., when the speaker perceives  
12 her own productions). In addition, they can receive activation from concepts and lemmas. The  
13 discrepancy between the expected activation (propagated in the form of inhibitory connections from  
14 the motor targets) and the actual activation constitutes a prediction error, which is propagated back  
15 to the motor target areas and used for online corrections of motor programs (and learning of more  
16 accurate motor-to-sensory mappings).

17 The inhibitory connections therefore implement a form of prediction that maps from motor  
18 commands to expected sensory consequences of executing those commands. The excitatory  
19 backward connections, instead, implement a form of inverse correction, which maps from sensory  
20 prediction errors to changes in the motor commands needed to compensate for those errors.  
21 Evidence for a fast-cycling loop at the phonetic level, in which motor representations are rapidly  
22 mapped onto sensory representations and *vice versa*, comes from several demonstrations that  
23 speakers compensate very quickly for perturbed auditory feedback (e.g., Houde & Jordan, 1998;  
24 Jones & Munhall, 2002; Tourville, Reilly, & Guenther, 2008).

25 The HSFC model is closely related to the DIVA/GODIVA model proposed by Guenther and  
26 colleagues (see Tourville & Guenther, 2011). This model also incorporates the notion that  
27 somatosensory and auditory target areas are activated via forward-model predictions as well as via  
28 processing of sensory input, and that prediction errors are used for online correction (as well as for  
29 learning the mappings between movements and their sensory consequences). Importantly, both  
30 models incorporate what we might call an account of self-monitoring. They assume that a process  
31 that maps from motor areas to sensory areas (and *vice versa*) is essential for online control during  
32 speech production (see Plaut & Kello, 1999 for another computational model that instantiates this  
33 idea).

Crucially, unlike psycholinguistic theories of self-monitoring (see Sections 3.2 and 3.3), these neuro-computational models focus on sound-level representations and processes, and say very little about other linguistic levels. Hickok (2012, 2014) argued for the importance of integrating psycholinguistic theories of language production with models of speech production, and integrated lemma and conceptual representations within his HSFC. However, he did not explicitly extend the forward-model architecture to these levels. In terms of Figure 1B, his model assumes that dotted arrows flow in both directions between all levels, indicating that both comprehension and production processes take place during language production. But the HSFC model includes a fast-cycling within-level loop (recursive dotted arrows), which is responsible for fast error correction during production, at the phonological and phonetic levels only. No such loop is explicitly assumed at the lemma and conceptual levels.

#### *An early model linking production and comprehension: the Perceptual Loop Theory of self-monitoring*

According to the Perceptual Loop Theory (Levelt, 1983; 1989), production errors are detected via comprehension of speech output (the external loop), and also via comprehension of phonological representations (the internal loop). The comparison process takes place at the level of communicative intentions (messages): If the message reconstructed by the comprehension system does not match the message originally intended, the monitor flags up an error. This comprehension-production loop is depicted in Figure 1B using a solid black line. This loop links comprehension representations at the phonological (phonological comprehension representations) and phonetic level (the speech percepts formed during comprehension of the speech output) to semantic representations (that are shared between comprehension and production).

Crucially, the Perceptual Loop Theory is open to criticism because it posits a relatively slow-cycling loop. First, the speech signal must be analyzed by comprehension processes to recover speech percepts (if using the external loop), or phonological representations retrieved by production processes have to be analyzed by comprehension processes to activate corresponding phonological representations in the comprehension network (if using the internal loop). Additional comprehension processes then map from sound-based comprehension representations to a semantic comprehension representation. Finally, the activated semantic representation in the comprehension network is compared to the semantic representation that was originally activated in the production network (the latter process is facilitated by shared representations at the semantic level). In addition,

1 if a discrepancy is detected, production of the current utterance must be stopped before production  
2 of a replacement can start.

3 Oomen and Poostma (2002) found that limiting participants' processing resources by having  
4 them engage in a concurrent task while speaking caused them to stop speaking more quickly after  
5 the onset of an error (Oomen & Postma, 2002). If the time it takes to stop were attributable to a  
6 comprehension-based loop, then one would have expected that drawing attention away from the  
7 speech signal (as in a dual task condition) would have led to longer, not shorter stopping times.

8 Hartsuiker and Kolk (2001) criticized the Perceptual Loop Theory assumption that  
9 production of the erroneous utterance must stop before planning of the replacement begins. Instead,  
10 they proposed that stopping the current utterance and preparing a replacement can proceed in  
11 parallel. Indeed, speakers can often resume very quickly following an interruption (in less than  
12 100ms; Blackmer & Mitton, 1991), which suggests that they start planning the replacements before  
13 they stop articulation. More direct evidence comes from Hartsuiker, Catchpole, de Jong, and  
14 Pickering (2008; see also Gambi, Cop, & Pickering, in press), who showed that the time it takes to  
15 stop a word depends on how difficult it is to prepare a replacement word.

16 However, Hartsuiker and Kolk (2001) also assume that the monitor needs to detect an error  
17 in the phonological representation before sending a signal to stop production. Interestingly, there  
18 are also psycholinguistic theories of self-monitoring that posit a purely production-based monitor  
19 (see Postma, 2000 for discussion). For example, Nozari, Dell, and Schwartz (2011) proposed that  
20 error detection is based on the amount of noise associated with production processes. One argument  
21 in favor of production-based accounts is that they allow for very rapid error detection at all levels of  
22 the linguistic hierarchy. Below, we introduce Pickering and Garrod's (2013; 2014) comprehension-  
23 based theory of self-monitoring theory which addresses this issue by allowing the monitor to  
24 compare expected and actual comprehension representations at all linguistic levels, as soon as these  
25 representations become available.

26

27 *Prediction during production and comprehension: the integrated theory of language production*  
28 *and comprehension*

29 Pickering and Garrod (2013) described an integrated theory of language production and  
30 comprehension that is based on the notion of forward models. This notion of internal models is  
31 derived from the motor control literature (e.g., Wolpert, 1997) and is also part of the models

1 reviewed in *Neuro-computational models of speech motor control*. Crucially, Pickering and Garrod  
2 generalize it, assuming that forward models are computed at all levels of the linguistic hierarchy,  
3 and that they are involved not only in speech production but also in speech comprehension. Below,  
4 we first describe how forward models are implicated in self-monitoring, and then how they are  
5 implicated in prediction during comprehension.

6         During an act of production, the speaker forms a communicative intention (production  
7 command), which corresponds to the pre-linguistic message that the speaker intends to convey. The  
8 production command is sent to the production system (or, in Pickering and Garrod's terminology,  
9 the production implementer), and it triggers the retrieval of a set of production representations  
10 (semantics, syntax, and phonology). For example, if a speaker sees a kite and forms the intention to  
11 name this object, production processes would cause the retrieval of the corresponding concept  
12 (KITE), lemma (*kite*) and phonological form (/kaɪt/). Importantly, it takes several hundred  
13 milliseconds to retrieve such representations (see Indefrey & Levelt, 2004, for estimates). Once  
14 production representations have been retrieved, they can be processed by the comprehension system  
15 (or, in Pickering and Garrod's terminology, the comprehension implementer)<sup>i</sup>. Crucially, the theory  
16 assumes that the comprehension implementer has immediate access to production representations at  
17 all levels; so, for example, the semantic representation retrieved during production can be  
18 immediately comprehended, even before a phonological representation is built. In this respect, the  
19 proposal differs from the Perceptual Loop Theory of self-monitoring (Levelt, 1983).

20         In addition to retrieval of representations within the production and comprehension  
21 implementers, during an act of production a copy of the production command is sent to a forward  
22 model, which maps from the production command to the predicted comprehension representations  
23 that are about to be retrieved as a consequence of executing that production command. To return to  
24 our example of a speaker intending to name the picture of a kite, a forward model of this process  
25 could compute a prediction of aspects of the semantics (it's a flyable object), of the syntax (it's a  
26 noun)<sup>ii</sup>, and the phonology (it starts with a consonant), *before* the corresponding production  
27 representations are retrieved from memory. Therefore, predicted representations are typically ready  
28 before actual (implemented) representations. The process of self-monitoring constitutes the  
29 comparison between predicted and actual comprehension representations within the comparator (at  
30 any linguistic level). The resulting difference (the prediction error in motor control terms; Wolpert,  
31 1997) can be used to drive online corrections (and learning), just as it can in the models described in  
32 section *Neuro-computational models of speech motor control*.

1 In sum, the account of self-monitoring proposed by Pickering and Garrod (2013; 2014)  
2 differs from the Perceptual Loop Theory (Levelt, 1983) in that it posits loops between production  
3 and comprehension at all levels of the linguistic hierarchy, and both within and between levels (not  
4 just from phonetics and phonology to semantics; see dotted arrows in Figure 1B). Moreover, such  
5 loops are faster than the loops assumed by the Perceptual Loop Theory, because they are based on  
6 comparisons between predicted and actual comprehension representations, with predicted  
7 comprehension representations being the outcome of production processes (i.e., left-to-right dotted  
8 arrows in Figure 1B).

9 With regard to acts of comprehension, Pickering and Garrod (2007) proposed that the collection of  
10 cognitive mechanisms underlying prediction during language comprehension coincides with the  
11 language production system. Federmeier (2007) made a similar proposal based on evidence that the  
12 left hemisphere is more sensitive to predictability of upcoming words than the right hemisphere  
13 (and the neural substrate for language production is predominantly left-lateralized in Broca's area),  
14 and Dell and Chang (2014) have recently reinstated this idea as the core principle of their P-chain  
15 framework (see *The P-chain framework*, this chapter).

16 An earlier proposal (Kempen, 2000, 2014) argued that grammatical encoding (production)  
17 and grammatical decoding (comprehension) are performed by the same processing architecture.  
18 Parallels between sentence comprehension and sentence production (e.g., similar patterns of errors  
19 occur during subject-verb agreement in both production and comprehension; Bock & Miller, 1991;  
20 Pearlmuter, Garnsey, & Bock, 1999), as well as evidence from structural priming from  
21 comprehension to production (Bock, et al., 2007; Branigan, et al., 2000) are consistent with this  
22 proposal. However, such evidence is indirect and, can be explained by shared representations  
23 without shared processes. In addition, it is clear how production and comprehension could share  
24 some processes (e.g., retrieving syntactic frames) but some differences in processes are necessary in  
25 order to explain the different start- and end-points. Note that the proposal that production processes  
26 are related to only a subset of comprehension processes, namely those involved in prediction during  
27 acts of comprehension, is not subject to this criticism.

28 But is there evidence for this hypothesis? Some evidence that production processes might be  
29 used during syntactic aspects of comprehension comes from a study by Kempen, Olsthoorn, and  
30 Sprenger (2012). They asked Dutch participants to paraphrase sentences from direct (e.g., *De*  
31 *lottowinnaar/ zei: / "Ik / heb besloten / een rode auto / te kopen / voor mezelf"*, The lottery winner  
32 said: "I have decided to buy a red car for myself") into indirect speech (e.g., *De lottowinnaar zei dat*  
33 *hij had besloten een rode auto te kopen voor zichzelf*, The lottery winner said that he had decided to

buy a red car for himself) as they read them (i.e., fragment by fragment, as marked in the example). When the sentence contained an ungrammatical reflexive pronoun (e.g., the third-person reflexive pronoun in the sentence *De lottowinnaar zei: “Ik heb besloten een rode auto te kopen voor zichzelf”*), participants were faster producing the paraphrase (which contained the same third-person pronoun) than when the sentence contained a grammatical reflexive (i.e., *mezelf*), despite the fact that the ungrammaticality should have led to a processing delay in comprehension. One interpretation of these findings is that participants’ expectations generated during the comprehension of the input sentences were replaced, on-line, by the expectations generated during concurrent encoding of the paraphrase (in which the same pronoun was grammatical).

Moreover, Federmeier, Kutas, and Schul (2010) showed that a late prefrontal positivity induced by plausible but unexpected nouns (which is thought to index the updating of disconfirmed predictions after an unexpected word has been encountered; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007) is greatly reduced in older adults (compared to younger adults) and, importantly, the magnitude of this component in the older group correlated with production measures of verbal fluency. Similarly, Mani and Huettig (2012) found that 2-year-olds with larger production (but not comprehension) vocabularies were more likely to predict upcoming referents (as indexed by looks to corresponding pictures in the so-called visual world paradigm; cf. Altmann & Kamide, 1999). More recently, a similar correlation between verbal fluency and prediction abilities during language comprehension (again, measured using the visual world paradigm) was reported for young adults as well, but only when listeners could preview pictures in the visual display (and presumably started retrieving their names; Hintz, Meyer, & Huettig, 2014). These studies suggest that the ability to predict during language comprehension is correlated with language production abilities at least in some task contexts.

In accordance with this and related evidence, Pickering and Garrod (2013) proposed that production processes underline comprehenders’ ability to predict what another is about to produce. This route to comprehension is termed prediction-by-simulation. It starts with the comprehender covertly imitating the producer: this means that, based on the initial part of the producer’s utterance, or on contextual information (i.e., what he assumes about the producer from previous interactions, or from background knowledge), the comprehender recovers the most likely intention (production command) underlying the utterance at time *t*. He can then run this command through the production implementer. If he does that, he will end up imitating the producer. This mechanism therefore explains alignment (Pickering & Garrod, 2004; see *Representational parity*).

1           In addition, the comprehender can run ahead the command that he recovered at time  $t$ , thus  
2 predicting what he would be likely to utter next if he were in the producer's shoes. If he runs this  
3 new command through the production implementer, he will be able to complete the producer's  
4 utterance (see *Dialogue*). If he runs this new command through the forward production model, the  
5 comprehender may generate the predicted semantics, syntax, and phonology at time  $t+1$ . When the  
6 producer continues his utterance, the comprehender builds comprehension representations for the  
7 actual utterance at  $t+1$  and can compare them to the representations he had predicted. He can then  
8 use the resulting discrepancy to adjust the recovered production command, thus revising his  
9 understanding of the intention underlying the producer's utterance.

10           In addition to the correlational evidence cited above, one study established a causal link  
11 between production processes and prediction during language comprehension (Lesage, Morgan,  
12 Olson, Meyer, & Miall, 2012). This study applied repetitive Transcranial Magnetic Stimulation  
13 (rTMS) to the right cerebellum. In rTMS, as in other types of TMS, a magnetic coil is used to  
14 induce small electric currents in a particular area of the brain; in particular, several pulses are  
15 delivered at a low-frequency, which is known to suppress neural activity for some time after  
16 stimulation has ended. Importantly, evidence suggests that forward model computations related to  
17 motor execution take place in the cerebellum (e.g., Wolpert, Miall, & Kawato, 1998), and some  
18 have linked it to the computation of internal models in general (e.g., Ito, 2008). Disrupting activity  
19 in the right cerebellum caused participants in this study to delay their eye-movements to predictable  
20 visual referents during sentence comprehension (but not to unpredictable visual referents).  
21 Crucially, other conditions with no stimulation or stimulation to a control site did not show the  
22 same selective effect. Therefore, this study suggests that forward model computations might  
23 support prediction during comprehension.

24           Note that, because the forward model is functionally distinct from the production  
25 implementer, Pickering and Garrod's (2013) account does not claim that full activation of the  
26 production implementer will be observed whenever prediction-by-simulation is used. For example,  
27 the account does not predict that activation of language production areas in the brain will be always  
28 observed during language comprehension. Rather, such activation is most likely to occur under  
29 conditions in which prediction-by-simulation is relied upon more. According to Pickering and  
30 Garrod, this is the case when comprehension is difficult (e.g., in a noisy environment; see Adank,  
31 2012).

32           But in addition, there is another route to prediction available in comprehension, which they  
33 termed prediction-by-association. This route does not involve production processes (i.e., forward



models or the production implementer), and could be used whenever covert imitation fails. This route to prediction in comprehension makes use of regularities in the input to the process of comprehension. Unlike prediction-by-simulation, it does not rely on knowledge of how we produce language. Instead, it relies on our ability to learn regular patterns of perceptual events, which applies equally to domains in which we have the ability to generate the patterns through action and domains in which we lack this ability (e.g., predicting the sound of leaves moved by the wind).

In sum, Pickering and Garrod's (2013) account allows for the existence of processes that are not (usually) shared between acts of comprehension and acts of production. In other words, there are some production processes that do not always operate during acts of comprehension (e.g., processes involved in retrieving articulatory programs; prediction-by-simulation via forward model computations), and there are comprehension processes that do not operate during acts of production (e.g., the prediction-by-association route).

#### *The P-chain framework*

The P-chain framework (Dell & Chang, 2014) claims that the process responsible for prediction during acts of comprehension is the same process that is used during acts of production. Therefore, although it does not claim that all processes are shared between comprehension and production, it posits that there is a single cognitive architecture subserving both tasks.

This assumption stems from the architecture of Chang, Dell, and Bock's (2006) Dual Path model of sentence acquisition (in children) and structural priming (in children and adults). This model is a recurrent neural network; during training, the model learns to predict the next word in the input (as in Elman, 1990). Prediction errors are generated by comparing the predicted with the actual comprehended input, and are used to change the weights between units in the network, so that the model learns to correctly predict grammatical word sequences in the future. Sometimes the model uses meaning (inferred from context) to help in this prediction process and the ability to do prediction from meaning is the same mechanism that the model uses for production. Thus, prediction from meaning (production) and prediction without meaning (comprehending other's utterances) involve the same mechanisms<sup>iii</sup>.

Dell and Chang (2014) proposed that the model's architecture instantiates a set of principles that govern the functioning of the cognitive system for language. They termed this collection of principles the *P-chain framework*. The framework highlights the tight links existing between

1 language comprehension (which they term *processing*) and *production*. Such links are organized in  
2 a chain, or loop, of cause-effect relations. In a nutshell, the use of production-based prediction  
3 mechanisms during language comprehension generates prediction errors, which in turn drive  
4 changes in the language system (i.e., they make it more likely to generate predictions in line with  
5 previous input), thus providing an explanation for structural priming effects, that is for the fact that  
6 comprehending a given sentence structure makes it more likely that the same structure will be  
7 selected in a subsequent act of production. When these changes build up over time, they can explain  
8 acquisition of structural representations for different languages (English: Twomey, Chang, &  
9 Ambridge, 2014; Japanese: Chang, 2009; German: Chang, Baumann, Pappert, & Fitz, in press).  
10 Finally, input regularities, on which comprehension is tuned, are themselves the output of the  
11 mechanism responsible for language production (in other speakers); therefore, there is also a slow-  
12 cycling loop through which the production processes of other speakers provide the input that trains  
13 comprehension processes. This final link is inspired by the Production-Distribution-Comprehension  
14 account (MacDonald, 2013) that is examined in the next section.

15

#### 16 *Frameworks that posit linked preferences*

17 Several theorists have appealed to the idea that regular patterns in language use emerge as a  
18 consequence of the constraints imposed by communication, and that such patterns in turn affect  
19 how speakers and listeners process language. This approach corresponds to positing a long-term  
20 loop (dashed arrow in Figure 1B) that is external to the language production and comprehension  
21 architecture (i.e., outside the mind/brain of individual language users).

22 One version of this idea is that speaker choices in production are constrained by ease-of-  
23 comprehension principles. For example, the Hyper- And Hypo-Articulation model of speech  
24 production (Lindblom, 1990) claims that speakers' tendency to reduce articulation (and therefore  
25 their own effort) is constrained by the necessity for listeners to recover the intended message. This  
26 leads speakers to counteract the tendency to hypo-articulate (i.e., to produce forms that are reduced  
27 in duration and/or intensity), precisely in those contexts in which the listener cannot draw on other  
28 sources of information (i.e., other than the speech signal) to infer meaning. There are other versions  
29 of this idea in phonetics (e.g., the Smooth Signal Redundancy hypothesis, Aylett & Turk, 2004).

30 Similarly, the Uniform Information Density hypothesis (UID; Levy & Jaeger, 2007; Jaeger,  
31 2010) claims that producers strive to keep information transfer rate within the range of the  
32 comprehender's processing rate (i.e., channel capacity); in this way, they avoid conveying too much

1 information or too little information per unit (word, phoneme, etc.). This can also be phrased in  
2 terms of surprisal, which is the predictability of a unit in context. High surprisal means that a unit is  
3 unpredictable based on previous context, and therefore adds information in that context, whereas  
4 low surprisal means that a unit is highly predictable and adds little information. UID claims that  
5 when given a choice speakers will tend to use a structure that keeps surprisal relatively constant  
6 across units. Production preferences are therefore explained as stemming from the limitations of the  
7 comprehension system. Interestingly, these preferences may be shaped by learning: Jaeger and  
8 Snider (2013) provide evidence that the higher the surprisal of a structural alternative in  
9 comprehension, the more likely that alternative is to be subsequently preferred in production.

10         The idea that production preferences can be explained with reference to comprehension is  
11 also related to the Audience Design Hypothesis: the notion that producers take into account their  
12 addressee's knowledge when planning their utterances (e.g., Brown-Schmidt, 2009; Clark, 1996). A  
13 review of this literature is beyond the scope of this chapter. However, we note that experimental  
14 evidence for the extent to which audience design affects production is mixed. In particular, there is  
15 controversy over the rapidity with which the addressee's knowledge can affect production. Keysar  
16 and colleagues have argued that producers are egocentric (e.g., Horton & Keysar, 1996), and take  
17 into account what their addressee can or cannot know only when given sufficient time and during a  
18 relatively late stage of production. In addition, it has been suggested that while speakers might  
19 adapt at the level of lexical choices, they in fact do not do so at the phonetic level (Bard et al., 2000;  
20 but see Galati & Brennan, 2010 for criticism, and Arnold, Kahn, & Pancani, 2012 for some  
21 evidence that speakers might adapt at least at the level of production speed). Finally, addressees are  
22 clearly facilitated when speakers adopt to previously established referential labels, but we do not  
23 know whether such facilitation occurs because listeners are sensitive to mutual knowledge between  
24 them and the speaker (Brown-Schmidt, 2009) or simply because repetition increases availability  
25 (Barr & Keysar, 2002).

26         Another related view is the Production-Distribution-Comprehension account (MacDonald,  
27 2013), which assumes that the relevant constraints shaping patterns of language use over time relate  
28 to production rather than to comprehension. In particular, it assumes three principles: (i) Easy First:  
29 Words that are more easily retrieved are produced first (e.g., Bock & Warren, 1985); (ii) Plan  
30 Reuse: Utterance plans that have been recently used tend to be reused (e.g., Bock, 1986); and (iii)  
31 Reduce Interference: Elements that are more similar to one another (and therefore tend to interfere  
32 in memory) are placed farther apart (Gennari, Mirković, & MacDonald, 2012 for semantic  
33 similarity; Jaeger, Furth, & Hilliard, 2012 for phonological similarity). The claim, then, is that these

constraints on production lead to distributional regularities to which comprehenders adapt as they accumulate linguistic experience. Therefore, linguistic forms that are easier to produce become easier to comprehend as well. This account thus explains both why certain structures are easier to comprehend than others and cross-linguistic patterns of language variation (i.e., typology). We refer the reader to Chapter 3 in this volume for an in-depth discussion of the evidence in favour and against the frameworks briefly introduced in this section.

## Dialogue

As mentioned in the Introduction, the nature of language use in dialogue (i.e., conversation) is one key motivation for positing links between comprehension and production. First, each participant in a dialogue regularly has to switch between acts of production and acts of comprehension. Such switches do not only occur between dialogue turns, but also within a turn, as listeners produce backchannels (e.g., *Yes*, *OK*, or *eh?*) to provide continuous feedback to the speaker. Moreover, such switches occur rapidly, as long intervals between turns are rare (Stivers et al., 2009). Finally, such switches can occur at any point within an utterance, with listeners taking over from speakers even after single words or incomplete constituents, and sometimes producing grammatical and pragmatically appropriate completions to these fragments (e.g., Clark & Wilkes-Gibbs, 1986; Lerner, 1991).

These phenomena demonstrate that the output of comprehension processes can rapidly affect production processes. For example, understanding the speaker's utterance leads to rapid backchannel responses from the listener and such backchannels can be quickly acted upon by the speaker. This suggests that loops must exist between comprehension processes and production processes, and that these loops must operate at a relatively fast rate. Moreover, the phenomenon of collaborative turn completions also suggests continuity between production and comprehension processes. Take the excerpt below (from Kurtić, Brown, & Wells, 2013, p. 726). B produces his utterance as a completion to the first part of A's utterance (*so I'm not sure whether they'll still be so willing to volunteer but I'll*) and it is so well-timed that it overlaps with the end of A's own turn (brackets indicate the start and end of speech produced in overlap by A and B). In turn, A shows evidence of having understood (and accepting) B's completion, despite it overlapping with her own turn; in fact, she goes as far as repeating B's utterance word by word.

A: ...so I'm not sure whether they'll still be so willing to volunteer but I'll [send them an email and ask]

1 B: [tell them about the free lunch]

2 A: I'll tell them about the free lunch

3 Finally, inter-turn intervals tend to cluster around a value that varies between 0 and 200ms (across  
4 languages; Stivers et al., 2009), with both long gaps and long overlaps being comparatively rare.  
5 This further suggests close links between production and comprehension, and has prompted the  
6 suggestion that comprehenders might be able to anticipate turn ends (de Ruiter, Mitterer, & Enfield,  
7 2006; Magyari & de Ruiter, 2012).

8

## 9 **Summary**

10 In this chapter, we have described several models that posit explicit links between  
11 production and comprehension. It is generally agreed that information about concepts, lemmas, and  
12 syntactic frames is shared between production and comprehension processes. But opinions diverge  
13 on the degree of sharing of phonological information, and most theorists assume that phonetic  
14 representations are separate (despite some dissent). Both the language production and the language  
15 comprehension literatures have internal debates about the directionality of processes, but such  
16 debates have not been framed as debates about the sharing of processes between production and  
17 comprehension until quite recently.

18 In Figure 1A, we assumed that any process that maps from “higher” linguistic levels  
19 (semantics) to “lower” linguistic levels (phonology) should be named a production process (left-to-  
20 right arrows), and that every process that maps in the inverse direction (right-to-left arrows) should  
21 be named a comprehension process. Based on this definition, we identified three accounts that  
22 assume comprehension processes take place during acts of production, in the form of self-  
23 monitoring: Levelt's (1989) Perceptual Loop theory (solid lines in Figure 1B), neuro-computational  
24 models of speech motor control (such as Hickok, 2012), and Pickering and Garrod's (2013)  
25 integrated account of language production and comprehension (dotted lines in Figure 1B). Such  
26 accounts differ in terms of the nature and speed of the loops they assume exist between production  
27 and comprehension.

28 We presented two accounts that assume that production processes take place during acts of  
29 comprehension to support prediction: Pickering and Garrod's integrated account, and the P-chain  
30 framework. In addition, we briefly described a number of frameworks that posit slower-cycling  
31 loops between production and comprehension; that is, loops that are mediated by long-term

experience with language and that can explain the development of linguistic preferences (dashed lines in Figure 1B). Overall, many researchers assume some degree of sharing of processes, but the range of views on this issue is far wider than on the issue of shared representations. Finally, we noted that the assumption of links between language production and language comprehension is also motivated by language use in dialogue. We believe that more explicit theorizing on the relations between production and comprehension processes, in both monologue and dialogue, would benefit the field.

## Acknowledgements

We would like to thank Franklin Chang and an anonymous reviewer for useful discussion.

## References

- Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language*, 122(1), 42-54.
- Alario, F.X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology*, 53A(3), 741-764.
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247-264.
- Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, 19(3), 505-512.
- Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31-56.
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42(1), 1-22.
- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46(2), 391-418.

- 1 Bates, E. (1993). Comprehension and production in early language development: Comments on  
2 Savage-Rumbaugh et al. *Monographs of the Society for Research in Child Development, Serial No.*  
3 *233, 58(3-4), 222-242.*
- 4 Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in  
5 spontaneous speech. *Cognition, 39(3), 173-194.*
- 6 Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from  
7 language comprehension to language production. *Cognition, 104(3), 437-458.*
- 8 Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin &*  
9 *Review, 3(4), 395-421.*
- 10 Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from  
11 language comprehension to language production. *Cognition, 104(3), 437-458.*
- 12 Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23(1), 45-93.*
- 13 Bock, K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence  
14 formulation. *Cognition, 21(1), 47-67.*
- 15 Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue.  
16 *Cognition, 75(2), B13-B25.*
- 17 Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal*  
18 *of Experimental Psychology: Learning, Memory, and Cognition, 22(6), 1482.*
- 19 Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents  
20 during interactive dialog. *Journal of Memory and Language, 61(2), 171-190.*
- 21 Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human  
22 cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the*  
23 *National Academy of Sciences, 110(7), 2653-2658.*
- 24 Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and  
25 accessibility effects in Japanese and English. *Journal of Memory and Language, 61(3), 374-397.*
- 26 Chang, F., Baumann, M., Pappert, S., & Fitz, H. (in press). Do lemmas speak German? A verb  
27 position effect in German structural priming, *Cognitive Science, doi:10.1111/cogs.12184*

- 1 Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234.
- 2 Clark, H. H. (1996). *Using language*. Cambridge University Press.
- 3 Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22(1), 1-  
4 39.
- 5 Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word  
6 production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 345.
- 7 D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in  
8 discriminating normal and degraded speech sounds. *Cortex*, 48(7), 882-887.
- 9 Dell, G.S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological*  
10 *Review*, 93 (3), 283–321.
- 11 Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to  
12 comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological*  
13 *Sciences*, 369(1634), 20120394.
- 14 De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A  
15 cognitive cornerstone of conversation. *Language*, 82(3), 515-535.
- 16 Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179-211.
- 17 Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent  
18 change in young adults. *The Journal of the Acoustical Society of America*, 121(6), 3814-3826.
- 19 Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically  
20 modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*,  
21 15(2), 399-402.
- 22 Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language  
23 comprehension. *Psychophysiology*, 44(4), 491-505.
- 24 Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use  
25 of prediction during language comprehension. *Brain and Language*, 115(3), 149-161.
- 26 Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of  
27 sentential constraint on word processing. *Brain Research*, 1146, 75-84.



- 1 Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in  
2 perception: Evidence from choice and simple response time tasks. *Journal of Memory and*  
3 *Language*, 49(3), 396-413.
- 4 Galantucci, B., Fowler, C. A., & Goldstein, L. (2009). Perceptuomotor compatibility effects in  
5 speech. *Attention, Perception, & Psychophysics*, 71(5), 1138-1149.
- 6 Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception  
7 reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.
- 8 Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the  
9 speaker, or for the addressee? *Journal of Memory and Language*, 62(1), 35-51.
- 10 Gambi, C., Cop, U., & Pickering, M. J. (in press). How do speakers coordinate? Evidence for  
11 prediction in a joint word-replacement task. *Cortex*, doi:10.1016/j.cortex.2014.09.009
- 12 Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and  
13 semantic co-ordination. *Cognition*, 27(2), 181-218.
- 14 Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*,  
15 8(1), 8-11.
- 16 Gennari, S. P., Mirković, J., & MacDonald, M. C. (2012). Animacy and competition in relative  
17 clause production: A cross-linguistic investigation. *Cognitive Psychology*, 65(2), 141-176.
- 18 Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological*  
19 *Review*, 105(2), 251-279.
- 20 Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production.  
21 *Journal of Neurolinguistics*, 25(5), 408-422.
- 22 Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational  
23 test of the perceptual loop theory. *Cognitive Psychology*, 42(2), 113-157.
- 24 Hartsuiker, R. J., Catchpole, C. M., de Jong, N. H., & Pickering, M. J. (2008). Concurrent  
25 processing of words and their replacements during speech. *Cognition*, 108(3), 601-607.
- 26 Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews*  
27 *Neuroscience*, 13(2), 135-145.

- 1 Hickok, G. (2014). Towards an integrated psycholinguistic, neurolinguistic, sensorimotor  
2 framework for speech production. *Language, Cognition and Neuroscience*, 29(1), 52-59.
- 3 Hintz, F., Meyer, A.S., & Huettig, F. (2014). The influence of verb-specific featural restrictions,  
4 word associations, and production-based mechanisms on language-mediated anticipatory eye-  
5 movements. *Proceedings of the 27th Annual CUNY conference on human sentence processing*.  
6 Ohio State University. Columbus, OH.
- 7 Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground?.  
8 *Cognition*, 59(1), 91-117.
- 9 Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*,  
10 279(5354), 1213-1216.
- 11 Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the  
12 auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*, 14(8), 1125-  
13 1138.
- 14 Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production  
15 components. *Cognition*, 92(1), 101-144.
- 16 Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews*  
17 *Neuroscience*, 9(4), 304-313.
- 18 Jarick, M., & Jones, J. A. (2008). Observation of static gestures influences speech production.  
19 *Experimental Brain Research*, 189(2), 221-228.
- 20 Jaeger, F.T. (2010). Redundancy and reduction: Speakers manage syntactic information density.  
21 *Cognitive Psychology*, 61(1), 23-62.
- 22 Jaeger, T. F., Furth, K., & Hilliard, C. (2012). Incremental phonological encoding during unscripted  
23 sentence production. *Frontiers in Psychology*, 3, doi:10.3389/fpsyg.2012.00481
- 24 Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation:  
25 Syntactic priming is affected by the prime's prediction error given both prior and recent experience.  
26 *Cognition*, 127(1), 57-83.
- 27 Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: studies of  
28 Mandarin tone production. *Journal of Phonetics*, 30(3), 303-320.

- 1    Kempen, G. (2000). Could grammatical encoding and grammatical decoding be subserved by the  
2    same processing module? *Behavioral and Brain Sciences*, 23, 38-39.
- 3    Kempen, G. (2014). Prolegomena to a neurocomputational architecture for human grammatical  
4    encoding and decoding. *Neuroinformatics*, 12, 111-142.
- 5    Kempen, G., Olsthoorn, N., & Sprenger, S. (2012). Grammatical workspace sharing during  
6    language production and language comprehension: Evidence from grammatical multitasking.  
7    *Language and Cognitive Processes*, 27, 345-380.
- 8    Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: evidence from stimulus  
9    response compatibility. *Journal of Experimental Psychology: Human Perception and Performance*,  
10    26(2), 634.
- 11   Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs  
12   in agrammatism: Implications for lexical organization. *Brain and Language*, 74(1), 1-25.
- 13   Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects,  
14   and speech processing. *Cognition*, 107(1), 54-81.
- 15   Kurtić, E., Brown, G. J., & Wells, B. (2013). Resources for turn competition in overlapping talk.  
16   *Speech Communication*, 55(5), 721-743.
- 17   Lerner, G. H. (1991). On the syntax of sentences-in-progress. *Language in Society*, 20(03), 441-  
18   458.
- 19   Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S., & Miall, R. C. (2012). Cerebellar rTMS  
20   disrupts predictive language processing. *Current Biology*, 22(18), R794-R795.
- 21   Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41-104.
- 22   Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- 23   Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive*  
24   *Psychology*, 14(1), 78-106.
- 25   Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production.  
26   *Behavioral and Brain Sciences*, 22(1), 1-38.

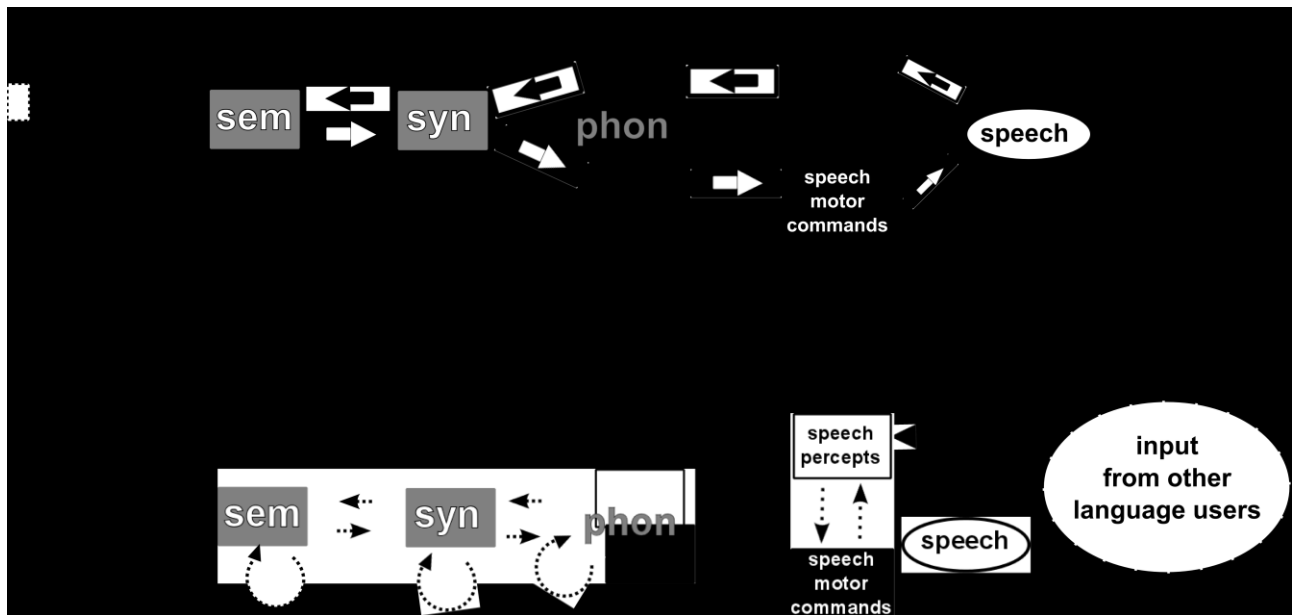
- 1 Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic  
2 reduction. In B. Schölkopf, J. Platt, and T. Hoffman (Eds.). *Advances in Neural Information*  
3 *Processing Systems*, Vol. 19 (pp. 849-856), Cambridge, MA: MIT Press.
- 4 Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in*  
5 *Cognitive Sciences*, 4(5), 187-196.
- 6 Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W.J.  
7 Hardcastle and A. Marchal (Eds.). *Speech Production and Speech Modelling* (pp. 403-439).  
8 Springer Netherlands.
- 9 MacDonald, M. C. (2013). How language production shapes language form and comprehension.  
10 *Frontiers in Psychology*, 4, doi:10.3389/fpsyg.2013.00226
- 11 MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other*  
12 *cognitive skills*. New York: Springer-Verlag.
- 13 MacKay, D. G., Abrams, L., & Pedroza, M. J. (1999). Aging on the input versus output side:  
14 Theoretical implications of age-linked asymmetries between detecting versus retrieving  
15 orthographic information. *Psychology and Aging*, 14(1), 3-17.
- 16 Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming  
17 words. *Frontiers in Psychology*, 3, doi:10.3389/fpsyg.2012.00376
- 18 Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only  
19 for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*,  
20 38(4), 843-847.
- 21 Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*,  
22 25(1), 71-102.
- 23 Menenti, L., Gierhan, S. M., Segaert, K., & Hagoort, P. (2011). Shared language overlap and  
24 segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI.  
25 *Psychological Science*, 22(9), 1173-1182.
- 26 Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is  
27 phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1), 168-173.

- 1 Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a  
2 loose perception–action coupling in speech. *Attention, Perception, & Psychophysics*, 75(3), 557-  
3 575.
- 4 Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy  
5 represent? Evidence from speech production. *The Journal of Neuroscience*, 33(41), 16110-16116.
- 6 Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection?  
7 A conflict-based account of monitoring in speech production. *Cognitive Psychology*, 63(1), 1-33.
- 8 Oomen, C. C., & Postma, A. (2002). Limitations in processing resources and speech monitoring.  
9 *Language and Cognitive Processes*, 17(2), 163-184.
- 10 Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the*  
11 *Acoustical Society of America*, 119(4), 2382-2393.
- 12 Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence  
13 comprehension. *Journal of Memory and Language*, 41(3), 427-456.
- 14 Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological*  
15 *Bulletin*, 134(3), 427-459.
- 16 Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral*  
17 *and Brain Sciences*, 27(02), 169-190.
- 18 Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions  
19 during comprehension?. *Trends in Cognitive Sciences*, 11(3), 105-110.
- 20 Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and  
21 comprehension. *Behavioral and Brain Sciences*, 36(04), 329-347.
- 22 Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models.  
23 *Frontiers in Human Neuroscience*, 8, doi:10.3389/fnhum.2014.00132
- 24 Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech  
25 comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.). *The*  
26 *Emergence of Language* (pp. 381-415). Mahwah, NJ: Taylor and Francis.
- 27 Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring  
28 models. *Cognition*, 77(2), 97-132.

- 1 Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis  
2 for language. *Nature Reviews Neuroscience*, 11(5), 351-360.
- 3 Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006).  
4 Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of*  
5 *Sciences*, 103(20), 7865-7870.
- 6 Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in  
7 language production: Picture-word interference studies. *Journal of Memory and Language*, 29(1),  
8 86-102.
- 9 Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—  
10 candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10(4),  
11 295-302.
- 12 Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in  
13 language production and language comprehension—an fMRI study. *Cerebral Cortex*, 22(7), 1662-  
14 1670.
- 15 Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: motor cortical  
16 activation during speech perception. *Neuroimage*, 25(1), 76-89.
- 17 Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing  
18 voices: how cortical areas supporting speech production mediate audiovisual speech perception.  
19 *Cerebral Cortex*, 17(10), 2387-2399.
- 20 Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C.  
21 (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National*  
22 *Academy of Sciences*, 106(26), 10587-10592.
- 23 Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation  
24 of dialect variants. *Journal of Memory and Language*, 60(4), 487-501.
- 25 Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition  
26 and production. *Language and Cognitive Processes*, 26(7), 952-981.
- 27 Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory  
28 feedback control of speech. *Neuroimage*, 39(3), 1429-1443.

- 1 Twomey, K., Chang, F., & Ambridge, B. (2014). Do as I say, not as I do: A lexical distributional  
2 account of English locative verb class acquisition. *Cognitive Psychology*, 73, 41-71
- 3 Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates  
4 motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701-702.
- 5 Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*,  
6 1(6), 209-216.
- 7 Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in*  
8 *Cognitive Sciences*, 2(9), 338-347.
- 9 Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huottilainen, M., Shtyrov, Y., & Mäkelä, J. P. (2014).  
10 Two distinct auditory-motor circuits for monitoring speech production as revealed by content-  
11 specific suppression of auditory cortex. *Cerebral Cortex*, doi: 10.1093/cercor/bht351.
- 12 Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in  
13 speech perception. *Proceedings of the National Academy of Sciences*, 107(2), 592-597
- 14 Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and  
15 memory. *Psychological bulletin*, 123(2), 162-185.

1 Figure 1 – Panel A shows production processes (white arrows on black background) and  
 2 comprehension processes (black arrows on white background); Panel B shows loops: the solid line  
 3 is Levelt's (1989) Perceptual Loop; the dotted arrows are Pickering and Garrod's (2013) fast loops  
 4 between and within levels; the dashed arrows are external loops that include long-term effects of  
 5 exposure to distributional regularities in linguistic input (see *The P-chain framework*, and  
 6 *Frameworks that posit linked preferences*).





---

<sup>i</sup> Note that the description provided by Pickering and Garrod (2013), and summarized here, appears to imply separate production and comprehension representations at all levels. In practice, the theory is consistent with the idea that the same representations are accessed during production as well as during comprehension. The two sets of representations they assume correspond to the output stage of production and comprehension processes respectively, but such processes may have access to the same pool of representations.

<sup>ii</sup> In the example, we focus on single word retrieval. However, Pickering and Garrod (2013) have also discussed this process in relation to constituent ordering (p. 339).

<sup>iii</sup> Note that within this single cognitive architecture, the Dual path model incorporates separate weights for the word-to-syntax (“comprehension”) and syntax-to-word (“production”) directions, but both of these representations are hypothesized to be used in both comprehension and production tasks.